MODEL BASED FACE RECONSTRUCTION FOR ANIMATION

WON-SOOK LEE, PREM KALRA, NADIA MAGNENAT THALMANN

MIRALab, CUI, University of Geneva, Geneva, Switzerland
E-mail : {wslee, kalra, thalmann}@cui.unige.ch

In this paper, we present a new semiautomatic method to reconstruct 3D facial model for animation from two orthogonal pictures taken from front and side views. The method is based on extracting the hair and face outlines and detecting interior features in the region of mouth, eyes, etc. We show how to use structured snakes for extracting the profile boundaries and facial features. Then DFFD, a deformation process, is used for modifying a predefined or generic 3D head model to produce the individualized head. Texture mapping based on cylindrical projection is employed using a composed image from the two images. The reconstructed 3D face can be animated in our facial animation system.

1 Introduction

3D modeling of human face has wide application from video conference to facial surgery simulation. Reconstruction of the face of a specific person, i.e. cloning a real person's face, has even more interest in today's virtual world. With the fast pace in computing, graphics, and networking technologies, VR communication is gaining more and more importance and popularity. To decrease transmission data size, it is necessary to reconstruct a realistic head in an efficient way. However, cloning a real person's face has practical limitations in the sense of time, simple equipment and realistic shape.

We present our approach to clone a real face from two orthogonal views, emphasizing accessibility for any body with low price equipment. We first detect predefined feature points on pictures and modify a 3D head model with animation structure with corresponding texture mapping. There are several approaches for feature detection in an interactive or automatic way. We compare those methods, discuss their drawback and introduce our semiautomatic method. Then we describe a way to modify a generic head smoothly using a deformation tool, Dirichlet Free Form Deformations (DFFD). The final model can then be used for animation.

We organize this paper as follows. We give a review in section 2 with classification for existing methods to obtain data points and geometrical structure of a face. In section 3, we describe our reconstruction method from two orthogonal pictures using interactive structured snakes with multiresolution approach for edge detection or Sobel operator with color blending function. The methods to modify a generic model and texture mapping using cylindrical projection are also provided. In section 4, the overview of our facial animation system is given. Section 5 illustrates several possible applications with cloned faces. Finally in section 6, conclusion is given.

2 Related Work

In this section, we review methods for shape reconstruction. First, we describe methods which have been used for restoring (acquiring) geometrical data considering only shape. Section 2.2 gives methods where reconstruction is done to acquire both shape and structure. With restored structure, the reconstruction model can be animated.

2.1 Shape Reconstruction

To get a detailed matched shape, we need time-consuming manual job, a sophisticated equipment, or complicated algorithm. Most of them need one more process to get structured shape for animation. In this section we focus on a few methods to get detailed range data for face.

Plaster Model Magnenat Thalmann et al. [16] used plaster models in real world and selected facets and vertices marking on the models which are photographed from various angles to be digitized. Here the reconstruction approach requires a mesh drawn on the face and is time consuming, but can obtain high resolution in any interested area.

Laser Scanning In range image vision system some sensors, such as scanners, yield range images. For each pixel of the image, the range to the visible surface of the objects in the scene is known. Therefore, spatial location is determined for a large number of points on this surface. An example of commercial 3D digitizer based on laser-light scanning, is Cyberware Color Digitizer^{TM}. Lee et al. [13] digitized facial geometry through the use of scanning range sensors. However, the approach based on 3D digitization requires special high-cost hardware and a powerful workstation.

Stripe Generator As an example of structured light camera range digitizer, a light striper with a camera and stripe pattern generator can be used for face reconstruction with relatively cheap equipment compared to laser scanners. Stripe pattern is projected on the 3D object surface and it is taken by a camera. With information of positions of projector and camera and stripe pattern, a 3D shape can be calculated. Proesmans et al. [20] shows a good dynamic 3D shape using a slide projector, by a frame-by-frame reconstruction of a video.

Lighting Switch Photometry Lighting Switch Photometry uses three or more light sources for computing normal vectors for extracting shapes of static objects [5] or a moving human face [21]. This method assumes that the reflectance map is lambertian. By Lighting Switch Photometry, the normal vector can be computed at the points where three incident light sources illuminate. It is difficult to compute the accurate normal vector at the point where the intensity of radiance is small, such as shadowed regions.

Stereoscopy A distance measurement method such as stereo can establish the correspondence at certain characteristic points. The method uses the geometric relation over stereo images to recover the surface depth. The method usually results in sparse spatial data. Fua and Leclerc [7] used it mainly in textured areas by weighting the stereo component most strongly for textured image areas and the shading component most strongly for texture-less areas.

2.2 Structured Shape Reconstruction

Most of the above methods concentrate on recovering a good shape, but the biggest drawback is that they provide only the shape without structured information. To get a structured shape for animation, most typical way is to modify an available generic model with structural information such that eyes, lips, nose, hair and so on. We classify methods using range data[1] and without using range data.

2.2.1 With Range Data

The plaster marking method by Magnenat Thalmann et al. [16] mentioned in previous section has structure for animation because each point has its own labeling corresponding to animation model. Except in this method, it is necessary to add a structural information to a set of 3D points to make the model suitable for animation.

---

[1] In this paper, we use the word 'range data' for a detailed shape data even though it produces some ambiguity. Intensity image system such as stereos scopic and shading is also included as 'range data' if they provide a detailed shape data as a result.

Warping Kernels Williams [23] reconstructed a head using Cyberware^{TM} digitizer and apply warpware to animate the model. A set of warping kernels is distributed around the face, each of which is a Hanning (cosine) window, scaled to 1.0 in the center, and diminishing smoothly to 0.0 at the edge.

Mesh Adaptation Starting with a structured facial mesh, Lee et al. [13] developed algorithms that automatically construct functional models of the heads of human subjects from laser-scanned range and reflection data. After getting the large arrays of data acquired by the scanner, they reduce it into a parsimonious geometric model of the face that can eventually be animated efficiently. They adapt a generic face mesh to the data. Once the mesh has been fitted by the feature based matching technique, the algorithm samples the range image at the location of the nodes of the face mesh to capture the facial geometry. The node positions also provide texture map coordinates that are used to map the full resolution color image onto the triangles.

2.2.2 Without Range Data

The approach based on 3D digitization to get a range data often requires special purpose high-cost hardware. So a common way of creating 3D objects is reconstruction from 2D information which is accessible at low price. Two commonly used methods are an interactive deformation method which modifies or generates a surface employing deformation, and a reconstruction method with feature points which modifies a generic model after feature detection.

Interactive deformation Magnenat Thalmann et al. [18] used an interactive tool to generate a polygon mesh surface for creating figures. The major operations performed include creation of primitives, selection, local deformations and global deformations. It is more tedious and time consuming. However, it may be only possible way to digitize a historical personage whose pictorial or other source is not available and is useful to invent new characters.

Reconstruction with feature points There are faster approaches to reconstruct a face shape from few pictures of a face [8, 1, 11]. In this method, a generic model in 3D is provided in advance, and a limited number of feature points are detected either automatically or interactively on the two (or more) orthogonal pictures, and the other points on the generic model are modified by a special function. Then 3D points are calculated by just combining several 2D coordinates.

Kurihara and Arai [11] used an interactive method to get a few points, and a Delaunay triangulation for the conformation of the face and texture mapping. The result seems nice, but a big drawback is that they use too few points to modify the generic model. So if the generic model has very different shape, the result may not be similar to the person and texture mapping may also not work well. To increase accuracy, one should increase input points for modification of generic model. Ip and Yin [8] have very similar approach to the one of Akimoto et al. [1]. These two approaches tried to detect feature points automatically using dynamic template matching or LMCT (Local Maximum-Curvature Tracking) checking concave and convex points on the side profile of a face and a very simple filtering method to get interior points. It was a trial for automation, but the method they use to detect points does not seem to be very robust. In addition LMCT was designed to calculate convex or concave points which works well only for Mongoloid looking people.

3 Our Approach

We present our approach to reconstruct a real face from two orthogonal views. In this reconstruction, only some points (so called feature points) are extracted. These are the most characteristic points to recognize people and can be detected from front and side views. It is a simple approach and does not require high cost equipment.

Reconstruction of a shape may not require high accuracy in some special cases, thanks to the texture mapping. However, for animation we need a process of texture fitting which can ensure positional correspondence of feature in the model and the texture image. The reliability of texture fitting is based on the number of feature points. Higher the number is, better the accuracy is. To make better reconstruction we need more points at the right place. If an automatic method is not robust enough, it is better to use interactive way. The overall flow for the reconstruction of 3D head is shown in Fig. 1. This is a part of an animation system described in section 4.

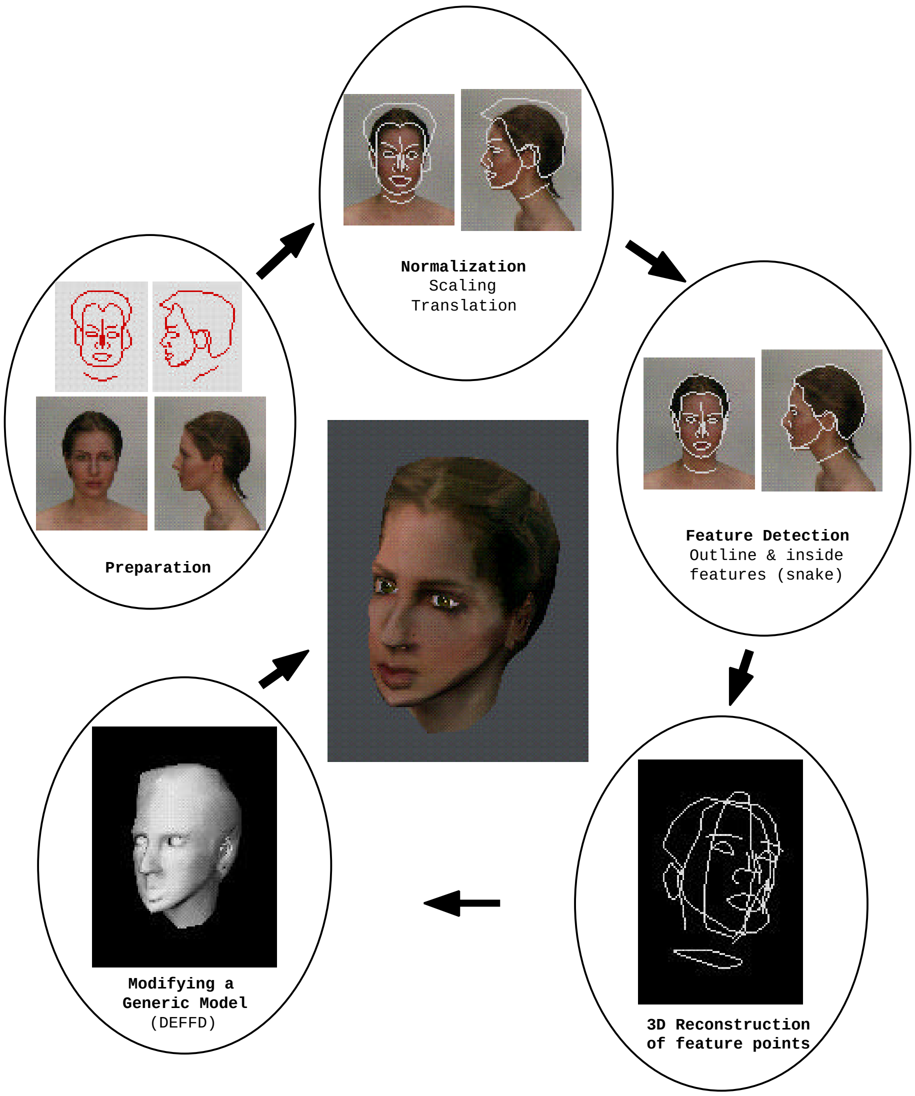

Figure 1: Overall flow for 3D head reconstruction

3.1 Preparation

First we prepare two 2D wire frames composed of feature points with predefined relation for front and side views. The frames are designed to be used as an initial position for the snake method later on. Then we take pictures from side and front views of a head. The picture is taken with maximum resolution and the face is in neutral expression and pose.

3.2 Normalization

To make the head heights of side and front views same, we measure heights of them, and choose one point from each view for matching them with corresponding points in prepared frame. As an example we select a highest point on a front face and a top nose on a side face. Then we use transformation (scaling and translation) to bring the pictures to the wire frame coordinate, overlaying frames on pictures.

3.3 Feature Detection

We provide automatic feature points extraction method with an interface for interactive correction if and when needed. We consider hair outline and face outline and some interior points such as eyes, nose, lips, eyebrows and ears as feature points. There are methods to detect them just using special background information and predefined threshold [8, 1] and then use an edge detection method and apply threshold again. Also image segmentation by clustering method is used [1]. However, it is not very reliable since the boundary between hair and face and chin line are not easy to detect in many cases. Moreover thresholding is too sensitive to each individual’s face image and therefore requires many trials and experiments. We use a structured snake for it which has functionality to keep the structure of contours. It does not depend on the background color much and is more robust than simple thresholding method. In addition it keeps the predefined number of points.

3.3.1 Structured Snake

First developed by Kass et al. [10], active contour method, so called as snakes, is widely used to fit a contour on a given image. This allows the fitting of the boundary points of maximum contrast close to the already-defined rough contour. To get correspondence between points from pictures and points on a generic model which has a defined number, a snake is a good candidate. Above the conventional snake, we add three more functions. First, we move few points to the corresponding position interactively, and anchor them. It helps later to keep the structure of points when snakes are involved and is also useful to get more reliable result when the edge we like to detect is not very strong. We then use color blending first for special area, so that it can be attracted by special color [3]. When the color is not very helpful and Sobel operator is not enough to get good edge detection, we use a multiresolution technique [4]. We can insert some more points which are not visible, but work as a member of snake to keep non uniform interval between visible points of the snake.

We have several parameters such as elasticity, rigidity, image potential and time step, to manage the movement of snake. As our contours are modeled as polylines, we use a discrete snake model with elastic, rigid forces and image force acting on each pixel for color interest. We define different sets of parameters for hair and face according to their color characteristic. To adjust the snake on points of strong contrast, we consider

$$F_{ext,\, i} = n_i \cdot \nabla E(v_i) \qquad (1)$$

where n_i is the normal to the curve at the node i, whose position is v_i and is given by

$$E(v_i) = | \nabla I(v_i) |^2 \qquad (2)$$

where $I(v_i)$ represents the image itself. To estimate the gradient of the image, we use the Sobel operator. Blending of the different color channels is changed to alter the snake's color channel sensitivity. For instance, we can make snakes sensitive to the excess of dark brown color for hair. Also we use clamping function to emphasize special interesting range of color. Snakes are useful in many circumstances, particularly in the presence of high contrast. The color blending and clamping function depend on the individual. See Fig. 2 for a result of hair boundary detection with some points anchored interactively.

Figure 2: The result for hair boundary detection after using snake without interactive correction. Some points with arrows are anchored to keep a structure.

Sobel operator does not always provide strong edge for some areas , for instance, chin lines. In such cases, we employ multiresolution approach [4] to obtain strong edges. It has two main operators, REDUCE with Gaussian operator and EXPAND. The subtraction produces an image resembling the result after Laplacian operators commonly used in the image processing. More times the REDUCE operator is applied stronger are the edges.

3.4 3D Reconstruction for feature points

We produce 3D points from two 2D points on frames with predefined relation between points on a front view and on a side view. Some points have x, y_f, y_s, z, so we take y_s , y_f or average of y_s and y_f for y coordinate (subscripts s and f mean side and front view). Some others have only x, y_f and others x, y_s. Using predefined relation from a typical face, we get 3D data x, y, z.

We modify non-feature points with some distance-related functions, spring-mass model, or FFD method. Here we employ Dirichlet Free Form Deformations (DFFD) to move other points according to feature points.

3.4.1 Modifying Generic Model

Distance-related functions have been employed by many researchers [8, 1, 11] to calculate displacement of non-feature points related to detected feature points in section 3.3. We propose to use DFFD [15] since it has capacity for non-linear deformations as opposed to generally applied linear interpolation which can give smooth result for the surface. The process of modification of a generic model fitting feature points detected from given two orthogonal pictures are as follows.

1. We define control points on a generic model which are corresponding to feature points detected from two views of a person and 27 points for a box surrounding.

2. Apply global transformations (translation, and scaling) to bring detected feature points to generic model's 3D space. We compare two eyes extremities in a generic model and a specific person for scaling. And we check the center of rightmost, leftmost, up-most, down-most, front-most, and back-most points of the head for translation.

3. Apply the DFFD on the points of the generic head. The displacement of non-feature points depends on the distance between control points. Since DFFD applies Voronoi and Delaunay triangulation, some points outside triangles of control points are not modified, the out-box of 27 points can be adjusted locally.

Our system provides a feedback modification of a head between feature detection and a resulted head. A modified model is shown in Fig. 3 with a generic head together.

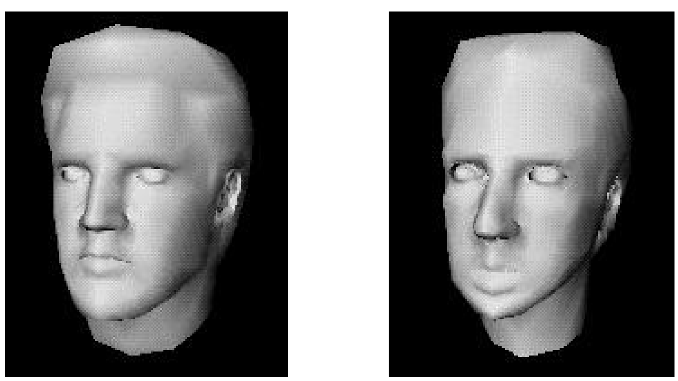

Figure 3: The result of DFFD modification comparing with the original head.

3.5 Texture Mapping

For smooth texture mapping, we assemble two images from front and side views to be one. First we use cylindrical projection of them. So the front view will cover from -90^0 to 90^0, right view from 0^0 to 180^0 and left one from -180^0 to 0^0. Since the front view is good only for a certain range and so is the other. To use conventional blending for wide range using angle variation [11], it is easy to blur certain shape of features. So we crop images for certain points (we use eye extremes because eyes are important to keep high resolution) on front and side views. Since we have information about eye positions, it can be done automatically to crop a front view, then we crop right and left views to make the final assembled image be 360^0. Since it is almost impossible to take two orthogonal pictures in exactly same condition, just to assemble three images makes the boundaries visible. We use a multiresolution spline method to assemble two images [4]. The cylindrical projections of front and right views and the image mosaic using a multiresolution spline are shown in Fig. 4.

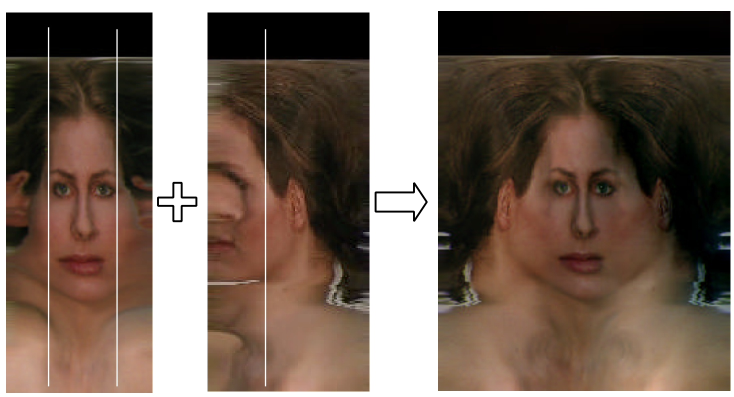

Figure 4: Front and right view after cylindrical projection covering 180^0 for each. We crop the front and side views around eye ends (shown through white lines) and combine with the left view (flipped from the right view). The last image shows image mosaic of three images, front, right and left using multiresolution spline method.

Texture mapping with a composed image is employed to 3D modified head. The main idea for the texture mapping is to map a 2D image to 3D shape. We lay projected 3D mesh of control (feature) points on the image and calculate Voronoi and Delaunay triangulation on 2D points. Then we calculate local barycentric coordinates of non feature points with a surrounding triangle of feature points. Then it determines texture coordinates on each point on a 3D surface from a 2D texture image [22]. Final textured image is shown in Fig. 5.

Figure 5: The final textured face.

4 Animation System

The (generic) face model is an irregular structure defined as a polygonal mesh. The face is decomposed into regions where muscular activity is simulated using rational free form deformations [9]. As model fitting transforms the generic face without changing the underlying structure the resulting new face can be animated. Animation can be controlled on several levels. On the lowest level we use a set of 65 minimal perceptible actions (MPAs) related to muscle movements.

The input can be in different form or media, and presently we concern two kinds, audio and video input. For audio input, we extract phoneme and use phoneme compiler which produce MPA arrays. For video input, we measure geometric features and use emotion compiler, then it also produces MPA arrays. We animate a reconstructed facial model with synchronized two MPA arrays [6]. Fig. 6 shows examples of facial animation on the reconstructed 3D face model.

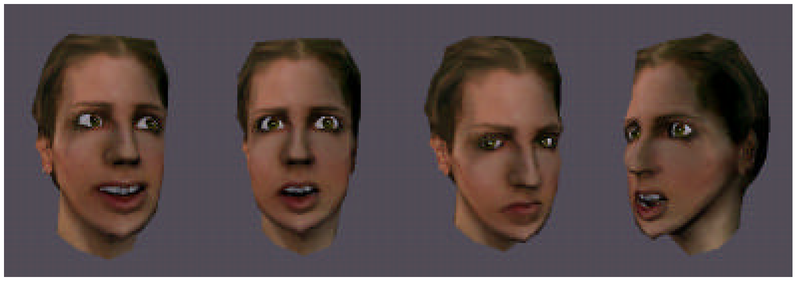

Figure 6: Several expressions on a reconstructed head.

5 Applications

In this section, we discuss about possible applications where reconstructed faces modified from a given generic model can be used.

5.1 Morphing

Face morphing has been also another interest for years. Magnenat-Thalmann and Thalmann use serial cross sections like tracing the contours from a topographic map [17] to morph between two animation models. There is a volume metamorphosis approach between CT human head and CT orangutan head [14]. However, more usual morphing between faces has been 2D morphing [12]. If we have reconstructed two faces modified from a generic model, it is easy to morph 3D face models since they have same number of points and structures.

5.2 Simulation of Generation

Genetic algorithm made a new idea to apply evolutionary theory into the computer, even to the virtual world making new generation from parents [2]. We can simulate new birth from parent if we have good models for parents. Also we can add generation factors which exist in new generation, something like most of new generation is taller, fatter and darker than older ones.

5.3 Communication

Recently virtual communication is getting more and more importance with faster network. People started to communicate on the net using teleconference, face by face. VR approach for communication have a lot of possibilities to extend and do dynamic behavior. In distributed VR with communication system [19] using model based coding for compression and transmissions for facial motion can employ the face reconstruction method to generate animated faces at remote site.

5.4 Medical Application

Many methods in vision have been developed to get face range data. If we have good shape with structural information, simulation of face surgery can have wider application since we can animate the face as we like to move.

6 Conclusion

We have developed a face reconstruction system from front and side views of a face using semi-automatic feature detection method. We use structured snake method to detect hair and face outlines and facial features. Then we use DFFD to modify a generic model and positional correspondence of feature points for texture fitting. With this method an animated facial model of a specific person can be easily created without needing any special and expensive equipment.

We plan to integrate this reconstruction method with facial expression recognition method. This way we will be able to extract both the shape and motion of the face.

Acknowledgment

The project is supported by Swiss National Research Foundation. The authors would like to thank other members of MIRALab for their help, particularly Philippe Gingins, Laurent Moccozet, Gael Sannier, and Laurence Suhner.

References

[1] Takaaki Akimoto, Yasuhito Suenaga, and Richard S. Wallace, Automatic Creation of 3D Facial Models, *IEEE Computer Graphics & Applications*, Sep., 1993.

[2] E. Baker and M. Seltzer, Evolving Line Drawings, In the *Proceedings of Graphics Interface,* pp. 91-100, 1994.

[3] P. Beylot, P. Gingins, P. Kalra, N. Magnenat Thalmann, W. Maurel, D. Thalmann, and F. Fasel, 3D Interactive Topological Modeling using Visible Human Dataset. *Computer Graphics Forum*, 15(3):33-44, 1996.

[4] Peter J. Burt and Edward H. Andelson, A Multiresolution Spline With Application to Image Mosaics,*ACM Transactions on Graphics*, 2(4):217-236, Oct., 1983.

[5] Coleman EN, Jain R, Obtaining 3-Dimensional Shape of Textured and Specular Surfaces Using Four-Source Photometry, *IEEE Computer Graphics and Image Processing*, 18:309-328, 1982.

[6] Marc Escher, N. Magnenat-Thalmann, Automatic 3D Cloning and Real-Time Animation of a Human Face, *Proc. Computer Animation*, IEEE Computer Society, pp. 58-66, 1997.

[7] P. Fua and Y.G. Leclerc, Taking Advantage of Image-Based and Geometry-Based Constraints to Recover 3-D Surfaces, *Computer Vision and Image Understanding*, 64(1):111-127, Jul., 1996.

[8] Horace H.S. Ip, Lijin Yin, Constructing a 3D individual head model from two orthogonal views. *The Visual Computer*, Springer-Verlag, 12:254-266, 1996.

[9] Prem Kalra, *An Interactive Multimodal Facial Animation System*, Ph.D. Thesis, Ecole Polytechnique Federale de Lausanne, 1993.

[10] M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active Contour Models, *International Journal of Computer Vision*, pp. 321-331, 1988.

[11] Tsuneya Kurihara and Kiyoshi Arai, A Transformation Method for Modeling and Animation of the Human Face from Photographs, *Computer Animation*, Springer-Verlag Tokyo, pp. 45-58, 1991.

[12] Seung-Yong Lee, Kyung-Yong Chwa, Sung Yong Shin, Image Metamorphosis Using Snakes and Free-Form Deformations, In *Computer Graphics (Proc. SIGGRAPH)* , pp. 439-448, 1995.

[13] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters, Realistic Modeling for Facial Animation, In *Computer Graphics (Proc. SIGGRAPH)*, pp. 55-62, 1996.

[14] Apostolos Lerios, Chase D. Garfinkle, Marc Levoy, Feature-Based Volume Metamorphosis, In *Computer Graphics (Proc. SIGGRAPH*), pp. 449-456, 1995.

[15] L. Moccozet, N. Magnenat-Thalmann, Dirichlet Free-Form Deformations and their Application to Hand Simulation, *Proc. Computer Animation*, IEEE Computer Society, pp. 93-102, 1997.

[16] N. Magnenat-Thalmann, D. Thalmann, The direction of Synthetic Actors in the film Rendez-vous à Montrèal, *IEEE Computer Graphics and Applications*, 7(12):9-19, 1987.

[17] N. Magnenat-Thalmann, D. Thalmann, *Synthetic actors in computer-generated 3D films*, Springer-Verlag, Tokyo. 1990.

[18] N. Magnenat-Thalmann, I. S. Pandzic, Jean-Claude Moussaly, Zhyong Huang, Jianhua Shen, The Making of Xian Terra-cotta Soldiers, *Computer Graphics: Developments in Virtual Environments*, (Academic Press) (ISBN 0-12-227741-4), 1995.

[19] Igor S. Pandzic, Tolga K. Capin, Nadia Magnenat Thalmann, Daniel Thalman Towards Natural Communication in Networked Collaborative Virtual Environments, In *Proceedings of FIVE '96*, Pisa, Italy, 1996.

[20] Marc Proesmans, Luc Van Gool. Reading between the lines - a method for extracting dynamic 3D with texture. In *Proceedings of VRST*, pp. 95-102, 1997.

[21] Hrroshi Saji, Hirohisa Hioki, Yoshihisa Shinagawa, Kensyu Yoshida and Tosiyasu L. Kunii, Extracting of 3D Shapes from the moving human face using lighting Switch Photometry. *Creating and Animating the virtual world*, Springer-Verlag, Tokyo, pp. 69-86, 1992.

[22] Sannier G, Magnenat-Thalmann N., A flexible texture fitting model for virtual clones, *Proceedings of Computer Graphics International*, IEEE Computer Society, pp. 167-176, 1997.

[23] Lance Williams, Performance-Driven Facial Animation, In *Computer Graphics (Proc. SIGGRAPH)*, pp. 235-242, 1990.